DOCUMENT RESUME

ED 339 212 FL 019 793

AUTHOR Abbott, Muriel
TITLE What Is LAB and Why Was It Renormed?
INSTITUTION New York City Board of Education, Brooklyn, NY. Office of Research, Evaluation, and Assessment.
PUB DATE Jun 91
NOTE 19p.
PUB TYPE Reports - Descriptive (141)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Communicative Competence (Languages); Elementary Secondary Education; English (Second Language); Language Proficiency; *Language Tests; *Spanish; *Standardized Tests; *Student Placement; *Test Norms; Test Reliability; Test Use; Test Validity
IDENTIFIERS *Language Assessment Battery

ABSTRACT

A report on the Language Assessment Battery (LAB) explains, in question-and-answer form, the causes and results of some changes made in the test norms. The LAB is a test of communicative language competence, written in English and Spanish versions and used for student placement in the New York City Public Schools. The report describes the test battery briefly and explains why the test of English language proficiency is given to non-native speakers of English, how scores are interpreted, how test norms are developed, why renorming was necessary, effects of the new norms, how renorming affected norms on the Spanish version, why the LAB is an appropriate measure of English language proficiency for students who are non-native speakers of English, and the LAB's reliability and validity. It is concluded that the renormed test battery reflects the same absolute level of language proficiency and also the change in norm group performance. The new norm-referenced scores do not reflect a decline in level of English language proficiency but merely a change in the basis of comparison. The introduction of the new norms will result in more limited-English-proficient students entitled to special services. Three sample cases are included. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED339212

OREA REPORT

WHAT IS LAB AND WHY WAS IT RENORMED?

June 1991

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Tobias, R.

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

New York City Public Schools

Joseph A. Fernandez
Chancellor

FL019793

BEST COPY AVAILABLE

NEW YORK CITY BOARD OF EDUCATION

Gwendolyn C. Baker
President

Irene H. Impellizzeri
Vice President

Carol A. Gresser
Westina L. Matthews
Michael J. Petrides
Luis O. Reyes
Ninfa Segarra
Members

Joseph A. Fernandez
Chancellor

ROBIN WILLNER
EXECUTIVE DIRECTOR
DIVISION OF STRATEGIC PLANNING/
RESEARCH AND DEVELOPMENT

It is the policy of the New York City Board of Education not to discriminate on the basis of race, color, creed, religion, national origin, age, handicapping condition, marital status, sexual orientation, or sex in its educational programs, activities, and employment policies, as required by law. Any person who believes he or she has been discriminated against should contact his or her Local Equal Opportunity Coordinator. Inquiries regarding compliance with appropriate laws may also be directed to Mercedes A. Nesfield, Director, Office of Equal Opportunity, 110 Livingston Street, Room 601, Brooklyn, New York 11201; or to the Director, Office for Civil Rights, United States Department of Education, 26 Federal Plaza, Room 33-130, New York, New York 10278.

7/3/90

ACKNOWLEDGEMENTS

This report was prepared by the Test Development Section of the Office of Research, Evaluation, and Assessment. The report was written by Muriel Abbott.

WHAT IS LAB AND WHY WAS IT RENORMED?

What is the Language Assessment Battery (LAB)?

LAB is a measure of language proficiency. There are English and Spanish versions. In LAB, language proficiency is defined as communicative competence; that is, the ability to convey and receive information through oral and written language. Within this context, LAB takes into consideration both academic and social language and aims at presenting tasks in the context of normal language usage. The ability to receive information is measured through the assessment of listening and reading skills, while the ability to convey information is measured through the assessment of speaking and writing skills. It is recognized that writing, as measured by LAB, is not a writing sample, but rather is a measure of elements of language usage that are essential to good writing: for example, correct use of parts of speech in context and recognition of good sentence construction.

Why is a test of English-language proficiency given to non-native speakers of English?

A test of English-language proficiency is given to non-native speakers of English for two purposes. The first purpose is for placement in appropriate instructional programs. It is important to identify those students whose level of English proficiency is such that they probably will not be successful academically without support services and who are, therefore, classified as limited-English-proficient (LEP). Students so identified are legally entitled to bilingual and English-as-a

second-language (ESL) instructional programs and support services that are more appropriate for their academic success. The second purpose is evaluation of the progress of entitled LEP students through these entitlement programs. The two purposes, individual LEP student placement and program evaluation, both require an instrument that measures differences in the level of English-language proficiency. Since LAB yields such measures, it can be used for both purposes.

How are scores on LAB interpreted?

Academic tests are concerned both with what a student knows and with how what he knows compares with what some defined group knows. LAB raw scores measure what a student knows; that is, how much language proficiency he has. That, of course, is of interest, but it is also necessary to know if that amount of language proficiency is enough for placement in a "non-entitlement class"; that is a "regular" all-English class without native language or ESL instructional support. For this purpose norm-referenced scores are used. In LAB the norm-referenced scores used are percentile ranks and normal curve equivalents (NCEs). The norm or comparison group, consists of native speakers of English because one wants to know how well a non-native English-speaking student's English-language proficiency compares with that of native-English-speaking students. A spring total test raw score of 61 in grade 3 reflects how much language proficiency a student has, but the corresponding percentile rank score of 16 indicates that based

on the 1989-90 norms, 16 percent of native speakers of English had spring total test raw scores at or below 61.

How are norms developed?

At each grade, a sample representative of the population of interest, such as native speakers of English, is selected. The test is administered to the sample. The scores obtained by students at each grade are then assembled into a cumulative frequency distribution so that the percent of students who scored at or below each raw score point can be identified.

Why was it necessary to renorm LAB?

The content of LAB has not changed. A total test raw score of 61 in spring of grade 3 still represents the same level of language proficiency. As norms age it becomes increasingly risky to base important decisions upon them. This is because the performance of the reference population tends to change over time. Therefore, norms need to be updated periodically. Since the early 1980's when LAB was normed, there has been improvement nationally in performance in language arts as reported by many test publishers. For example, two years ago, the Degrees of Reading Power (DRP) was renormed to reflect these changes in performance by the reference population (i.e. students in grades 2-10 nationwide). This situation exists also in the case of LAB. In order to know if a student's level of language proficiency is such that he probably will be successful without

bilingual/ESL services, his level of performance in language proficiency must be compared with that of the native speakers of English who are in the non-entitlement classes today. Until spring 1991, however, the student's current performance was being interpreted in terms of the norm group performance in 1981-82. This leads to inappropriate placement decisions. Placement decisions are more appropriately based on interpretation of a student's performance today in terms of up-to-date norms.

What is the effect of the new 1989-90 norms?

Again there is the issue of what or how much a student knows versus how what he knows compares with that of the norm group. Because the content of LAB has not changed, raw scores still represent the same absolute level of language proficiency. Only the performance of the norm group has changed. This however, affects the interpretation of the raw score. Because of the improved performance of the norm group the same amount of English language proficiency as reflected in the same raw score will result in a lower percentile rank score; a greater percentage of the 1989-90 norm group had scores above a particular raw score than did the 1981-82 norm group. This situation has implications both for the interpretation of an individual student's score and for the effect of this interpretation upon the number of students citywide who are identified as "LEP" and entitled to bilingual/ESL programs.

Originally the criterion for entitlement services was set by the Aspira Consent Decree at the total test raw score corresponding to the 20th percentile based upon the 1981-82 native-speakers-of-English norms. By 1988, largely as a result of the improved performance by native English speakers, this criterion resulted in students exiting from entitlement programs who all too often fail to perform successfully in non-entitlement classes. Also many new entrants into the New York City Public Schools who needed bilingual/ESL programs were not assigned to them. Therefore, beginning with the 1989-90 school year the New York City Board of Education mandated an upward revision in the entitlement criterion score to the total test raw score corresponding to the 40th percentile on the 1981-82 native speakers of English norms. This criterion was then applied in spring 1991 to the new 1989-90 norms, but the 40th percentile on the 1989-90 norms corresponds to a higher raw score than on the 1981-82 norms.

For individual students this means that although a student's level of English-language proficiency may be the same, it may no longer be sufficient to exit from an entitlement program. It also means that for some students whose level of English language proficiency formerly was between the 20th and 40th percentiles that level may now be below the 20th percentile.

Citywide, because the 40th percentile on the 1989-90 norms represents a higher level of performance, more students are identified as entitled to bilingual/ESL

services. The increased numbers of entitled LEP students resulting from the new norms is augmented by increased numbers of immigrants who are new entrants into the New York City Public Schools. Citywide, also, as a result of the new norms, there has been not only an increase in the total number of students identified as entitled, but also an increased proportion of those preforming below the 20th percentile while the proportion scoring between the 20th and 40th percentiles has decreased.

How did the renorming affect the norms on the Spanish version?

The Spanish version of LAB was designed to measure the Spanish language proficiency of native speakers of Spanish. It is used in New York City primarily to indicate language of dominance: Spanish or English. The Spanish version is not a translation of the English version but was developed concurrently with the English version and was designed to be comparable to it. The Spanish norms were based upon a selected sample of native-Spanish speakers in the New York City Public Schools. The native-speakers-of-Spanish norms are somewhat less difficult than are the native-speakers-of- English norms. Because of their exposure to an English-speaking environment, their Spanish was presumed to be somewhat less proficient than that of native speakers of English in New York City. The reverse of this situation would be expected in a country where Spanish is the native language.

Just as the English-language proficiency of native speakers of English improved from 1981-82 to 1989-90, so did the Spanish-language proficiency improve for native speakers of Spanish. Again, a particular raw score in 1989-90 still represents the same absolute level of Spanish-language proficiency as in 1989-90. However, because of the improved performance of the Spanish norm group, this same raw score results in a lower percentile rank score. In other words a greater percentage of the 1989-90 norm group had scores above that particular raw score than did the 1981-82 norm group. Just as it is important to interpret the English-language proficiency of today's students in terms of up-to-date native-speakers-of-English norms, so is it important to interpret the Spanish-language proficiency of today's students in terms of up-to-date native-speakers-of-Spanish norms.

Why is LAB an appropriate measure of English-language proficiency for students who are non-native speakers of English?

LAB was designed specifically for non-native speakers of English. Most measures of English-language proficiency have been designed for native English speakers. Because of this approach the difficulty of LAB is more appropriate for non-native English speakers. It was designed to be of average difficulty for these students with a within-level p-value of 50-60 for a fall administration. This means it has an appropriate range of difficulty for them. Appropriate difficulty level is conducive to more reliable measurement. Of course, a test of English language proficiency that is of average difficulty for non-native speakers of English will be very easy for native English speakers. This situation means that scores for limited-English-

proficient students are more normally distributed than are those for native-English speakers whose score distributions are very skewed: a piling up of high scores. This is not a problem because LAB is given to native-English speakers only for the purpose of developing norms. Other than for norms development, native-English speakers do not take LAB. The difficulty level of LAB or any test should be appropriate for those students who take it.

Is LAB a reliable instrument?

Reliability is a measure of the extent to which a test consistently measures whatever it is that it does measure. LAB is an extremely reliable test which means that the same results would be obtained with repeated test administrations. (Reliability coefficients (KR_{20}) are in the high .80s for individual subtests and in the .90s for total test.)

Is LAB a valid measure of English-language proficiency?

The validity of a test is specific to the purpose for which it is to be used and the group about whose performance one wishes to draw inferences. Therefore, there are different kinds of validity. In the case of LAB, content validity is crucial: that is, how well LAB samples from and reflects the objectives of relevant instructional programs. This was assured by reviewing, selecting and measuring curriculum objectives. An objective to test-item match reflects this correspondence.

Construct validity refers to how well a test reflects some underlying attribute of a student. In the case of LAB this is language proficiency. The possession of increased amounts of language proficiency should be reflected in higher scores. In LAB the within level grade-to-grade decreases in item difficulties reflects increased amounts of English-language proficiency as students progress through the instructional programs.

It is of utmost importance also that a measure of language proficiency perform in the same way for both limited- English-proficient students and native speakers of English if the performances of LEP students are to be interpreted in terms of the performance of a norm group of native speakers of English. This was supported by research that indicated that item difficulties rank order in the same way for both groups; the same items are easy or difficult for both groups.

SUMMARY

- A raw score in 1989-90 continues to reflect the same absolute level of language proficiency as that same raw score in 1981-82.

- The new 1989-90 norms reflect the change in the performance of the norm group. Norm-referenced scores are reported as percentile rank scores and Normal Curve Equivalents (NCEs).

- The drop in norm-referenced scores is the result of using the new norms, a basis of comparison which is tougher. These norm-referenced scores do not reflect a decline in the level of English-language proficiency of non-native speakers of English, merely that the basis of comparison has changed.

- The introduction of the new norms will result in an increase in the number of entitled LEP students since a higher raw score is required to reach the mandated total test 40th percentile on native-speakers-of-English norms.

Sample Case #1: Change within a level (Level I)

	Total Test Raw Score	Percentile Rank 1981-82	Percentile Rank 1989-90
Level I Grade 1 Spring 90	45	35	25
Level I Grade 2 Spring 91	50	26	19

In the example above, a grade 1 student in spring 1990 had a LAB total test raw score of 45. A raw score of 45 on Level I always represents the same level of English proficiency. However, for the raw score to have meaning, it must have a frame of reference such as the performance of a comparison group - the norm group.

Because test content is the same within a level and because the same level was given at both grades, this student's two raw scores can be compared directly to determine if a gain in proficiency occurred. For example, his LAB total test raw score of 50 in grade 2 can be compared directly with his 45 in grade 1. This comparison shows that he has gained in English proficiency by 5 raw score points. Based on the 1981-82 norms his percentile rank of 35 in grade 1 and 26 in grade 2 shows that his difference in proficiency (gain) between grades 1 and 2 was less than that of 1981-82 norm group. He did as well as or better than 35 percent of the 1981-82 norm group in grade 1 but better than only 26 percent of that group in grade 2. In other words, his raw score gain of 5 points was not sufficient to maintain his position with respect to the norm group. This situation is also true when his performance is interpreted in terms of the 1989-90 norm group.

His raw score in grade 1 of 45 had a percentile rank of 35 based on the 1981-82 norms and 25 on the 1989-90 norms. Similarly his grade 2 raw score of 50 had a percentile rank of 26 based on the 1981-82 norms and 19 based on the 1989-90 norms. This does not reflect a decline in his absolute level of language proficiency in either grade. The overall drop in percentile rank scores from the 1981-82 norms to the 1989-90 norms (35 to 25 in grade 1 and 26 to 19 in grade 2) is the result of the greatly improved performance of the 1989-90 norm group once that of the 1981-82 norm group.

Sample Case #2: Change within a level (Level IV)

	Total Test Raw Score	Percentile Rank 1981-82	Percentile Rank 1989-90
Level IV Grade 9 Spring 90	77	23	12
Level IV Grade 10 Spring 91	90	31	19

In the example above, a grade 9 student in spring 1990 had a LAB total test raw score of 77. A raw score of 77 on Level IV always represents the same level of English proficiency. However, for that raw score to have meaning, it must have a frame of reference such as the performance of a comparison group - the norm group.

Because test content is the same within a level and because the same level was given at both grades, this student's two raw scores can be compared directly to

determine if a gain in proficiency occurred. For example, his LAB total test raw score of 90 in grade 10 can be compared directly with his 77 in grade 9. This comparison shows that he has gained in English proficiency by 13 raw score points. Based on the 1981-82 norms his percentile rank of 23 in grade 9 and 31 in grade 10 shows that his 13 raw score point gain from grade 9 to grade 10 was greater than that of either norm group. In grade 9 his score of 77 was equal to or better than that of 23 percent of the 1981-82 group but in grade 10 his raw score of 90 was equal to or better than that of 31 percent of that group. This same improvement relative to the norm group is reflected in his percentile rank scores based on the 1989-90 norm group: 12 in grade 9 and 19 in grade 10.

His raw score of 77 in grade 9 had a percentile rank of 23 based on the 1981-82 norms and 12 on the 1989-90 norms. His raw score of 90 in grade 10 had a percentile rank of 31 based on the 1981-82 norms and 19 based on the 1989-90 norms. This does not reflect a decline in his absolute level of language proficiency. Whether one looks at his raw scores or his percentile ranks, his scores showed improvement in grade 10 over grade 9. The overall drop in his percentile rank scores from the 1981-82 to the 1989-90 norms (23 to 12 in grade 9 and 31 to 19 in grade 10) is the result of the greatly improved performance of the 1989-90 norm group over that of the 1981-82 group.

Sample Case #3: Change across levels (Level II to Level III)

	Total Test Raw Score	Percentile Rank 1981-82	1989-90
Level II Grade 5 Spring 90	75	26	16
Level III Grade 6 Spring 91	77	26	11

In the case above, a student in spring 1990 had a Level II grade 5 total test raw score of 75 and in spring 1991 had a Level III grade 6 total test raw score of 77. Because of the different test content at the two levels, the two raw scores cannot be compared directly. However, the Level III test was constructed to contain more difficult content than that of Level II. Therefore, by maintaining position in grades 5 and 6 at the 26th percentile on the 1981-82 norms, it can be assumed that the student showed a difference in proficiency comparable to that attained by the grade 5 and 6 students in the 1981-82 norm group. The introduction of the new 1989-90 norms somewhat complicates the interpretation. This student's level of proficiency as determined by his raw scores was at the 26th percentile in both grades 5 and 6 relative to the 1981-82 norm group. However, this same level of proficiency, as determined by his raw scores, was at the 16th and 11th percentiles relative to the 1989-90 norm group. It should be noted again that this does not necessarily mean a decline in the student's absolute level of proficiency.

It does mean that in both grades 5 and 6 the 1989-90 norm group performed better than did the 1981-82 norm group. Therefore, percentile ranks corresponding to the student's total test raw scores showed a decline. For example in grade 5, 26 percent of the 1981-82 norm group had a total test raw score below 75 whereas 16 percent of the 1989-90 norm group did so. In grade 6, 26 percent of the 1981-82 norm group had a total test raw score below 77 whereas 11 percent of the 1989-90 norm group did so. This also means that while in both grades 5 and 6 the 1989-90 norm group performed better than did the 1981-82 norm group, the 1989-90 norm group showed greater improvement in language proficiency over the 1981-82 norm group in grade 6 than in grade 5. Therefore, to be at the 16th percentile in grade 6 this student would have had to increase his total test raw score to 83. It should be stated again that the overall drop in his percentile rank scores from the 1981-82 to the 1989-90 norms (26 to 16 in grade 5 and 26 to 11 in grade 6) is the result of the greatly improved performance of the 1989-90 norm group over that of the 1981-82 group.